The impact of updating cancer survival methodologies for national estimates, 2019

Changes that have been made to the methodology for cancer survival in adults (aged 15 to 99 years)

Table of contents

1 . Introduction

This article describes changes to the bulletin [Cancer survival in England: adult, stage at diagnosis and childhood](#) which is jointly published by Office for National Statistics (ONS) and Public Health England (PHE). Multiple changes have been introduced into the method of calculating survival in England in adults and by stage at diagnosis for patients followed up to 2017. The impact of these changes is explained and a revised back series of data from patients diagnosed between 2006 to 2010 followed up to 2011 is provided.

There have been three changes to the methodology in producing cancer survival estimates. These are: adjustments to coding, enhancements to the cohort selection criteria and updated life tables. Section 2 details each change, providing the rationale for why the change has been made and what the change was. Section 3 outlines the impact of all changes on the [three most common cancers in England](#), which are breast, prostate and lung cancers.

The changes in methodology outlined in this publication will also affect the [Index of cancer survival for Clinical Commissioning Groups in England](#) and the [Geographic patterns of cancer survival in England bulletins](#). The back series for these publications will be published alongside the relevant releases later in the year.

2 . Changes to methodology

Adults' cancer survival estimates, overall and by stage or geographic level, are created by comparing cancer survival lengths with all-cause (or background) mortality. More information can be found in [Cancer survival statistical bulletins Quality and Methodology Information (QMI)](#). This section describes the changes to the methodology.

Update of life tables used to measure background mortality

Survival analysis involves comparing the survival of patients with cancer with background mortality (the survival of the general population). Background mortality is derived from population life tables, which are broken down by single year of age (0 to 99 years), sex, quintile of the Index of Multiple Deprivation (IMD) and region.

Previously, the life tables used were supplied by the [London School of Hygiene and Tropical Medicine's Cancer Survival Group](#) (LSHTM). English life tables for 1981 to 2011 were supplied, with denominators based on census data adjusted by estimates in migration. Previously, the 2011 life table was used for 2011 to 2015.

Due to recent [changing mortality trends](#), it is important for more up to date life tables to be used which take into account the slowdown in the long-term improvements to mortality rates. Office for National Statistics (ONS) has created these tables to use in calculating cancer survival. ONS currently [produces life tables](#) by single year of age at a national level and life tables by age group at a sub-national level. For cancer survival analysis, ONS have created life tables which look at [single year of age at a sub-national level](#), for 2001 to 2003 onwards. Smoothing has been applied to the life tables to overcome issues around high volatility and risk of disclosure. The method for producing these life tables is described in Annex 1.

The life tables supplied by LSHTM were used for calculating cancer survival estimates for publications up to and including cancer survival for 2005 to 2009 diagnoses followed-up to 2010. The life tables created by ONS will be used to calculate survival for 2006 to 2010 diagnoses followed-up to 2011 onwards.

Adjustments to coding

Three sets of changes that support the coding of datasets have been identified.

Correction to Income Domain of the Index of Multiple Deprivation

Cancer survival is calculated on an individual patient level and takes into account each person's age, sex, region and Income Domain of the Index of Multiple Deprivation (IDIMD) at diagnosis. It was identified that, in the follow-up to 2016 publication, there had been an error in the derivation of IDIMD quintiles, resulting in some people being assigned to the wrong quintile based on their postcode. This has been rectified for the follow up to 2017 release and the data has been corrected in the back series provided alongside this release.

Enhancement to the inclusion criteria

An important part of calculating cancer survival estimates is the criteria used to select the patients included in the cohort for analysis. Section 6 of the Cancer survival statistical bulletins QMI explains the criteria used to identify records that are eligible to be included in the calculation of cancer survival estimates.

The enhancements made build on the criteria used in the previous publication. The change made reflect the recommendations in the European Network of Cancer Registries' A proposal on cancer data quality checks: one common procedure for European cancer registries. They are the process for identifying site-specific inconsistencies in terms of sex and morphology, for example. More detail is provided in the Annex.

Correction to capturing dates of death

When a cohort of patients is checked to see if they are alive, dead or cannot be followed-up for vital status at a specified date (usually because they have emigrated from England), a very small proportion of patients are found to be dead but not yet notified to the cancer registry (NCRAS). These patients are now fully recognised in the data extracts.

These three sets of technical changes are discussed in combination under Tables 1 to 4 and are labelled as "adjustments to coding".

Enhancement to the cohort selection criteria

To produce estimates that are as robust and timely as possible for both 1- and 5-year cancer survival, the complete approach (Section 8 in the QMI) is used on 5-year cohorts. For example, 1-year and 5-year survival estimates for patients followed up to 2016 were based on patients diagnosed between 2011 and 2015.

For the Cancer survival in England: adult, stage at diagnosis and childhood – patients followed up to 2016 release, the back series was recreated using all data available at the time of analysis, in early 2017. For example, in the patients followed up to 2016 release, the 2006 to 2010 cohort included all information available in early 2017. This meant that the cohort included any late registrations and would have extra follow-up for some patients. The new method used means that the 2006 to 2010 cohort would not have any information available past early 2012, so would not include any late registrations or extra follow-up.

This change means that because cohorts are created consistently over time, estimates are comparable without the need to annually revise the back series.

3 . Impact of changes on cancer survival estimates

The reference tables provided show the revised back series estimates by cancer site, age and year of diagnosis. For the purpose of this summary we have focussed on the age-standardised survival of the three most common cancers in England, which are breast, prostate and lung.

Tables 1 to 4 present both previously published and the revised estimates (including all the changes mentioned in Section 2). Extra information is provided in the text on estimates with the adjustments to coding and estimates with both the adjustments to coding and revised cohort selection criteria (detailed results not supplied); this is to help understand the impact each change has had.

The tables also include information on significant changes between the published and revised estimates. Significance is determined by non-overlapping confidence intervals. Survival estimates are provided with 95% lower and upper confidence limits. These form a confidence interval, which is a measure of the statistical precision of an estimate and shows the range of uncertainty around the calculated estimate. As a general rule, if the confidence interval around one figure overlaps with the interval around another, we cannot say with certainty that there is more than a chance difference between the two figures.

Impact of changes on breast cancer survival estimates

Table 1: Impact of changes on age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with breast cancer, England

| Cohort | 1-year survival | | | | 5-year survival | | | |
	Published data (%)	Revised data (%)	Difference	Significant change?	Published data (%)	Revised data (%)	Difference	Significant change?
2006-2010	94.4	94.5	0.1	No	81.4	83.6	2.2	Yes
2007-2011	94.7	94.9	0.2	No	82.1	82.8	0.7	No
2008-2012	94.8	95.1	0.3	No	82.9	84.1	1.2	Yes
2009-2013	95.2	95.7	0.5	Yes	84.0	85.6	1.6	Yes
2010-2014	95.4	95.5	0.1	No	85.0	85.0	0.0	No
2011-2015	95.6	95.7	0.1	No	86.0	85.0	-1.0	Yes

Source: Office for National Statistics and Public Health England

For 1-year survival, the combination of changes has caused the survival estimate to increase for each cohort, with an average increase of 0.2 percentage points. The smallest increase was 0.1 percentage points and the largest increase was for the 2009 to 2013 cohort which saw a significant increase of 0.5 percentage points.
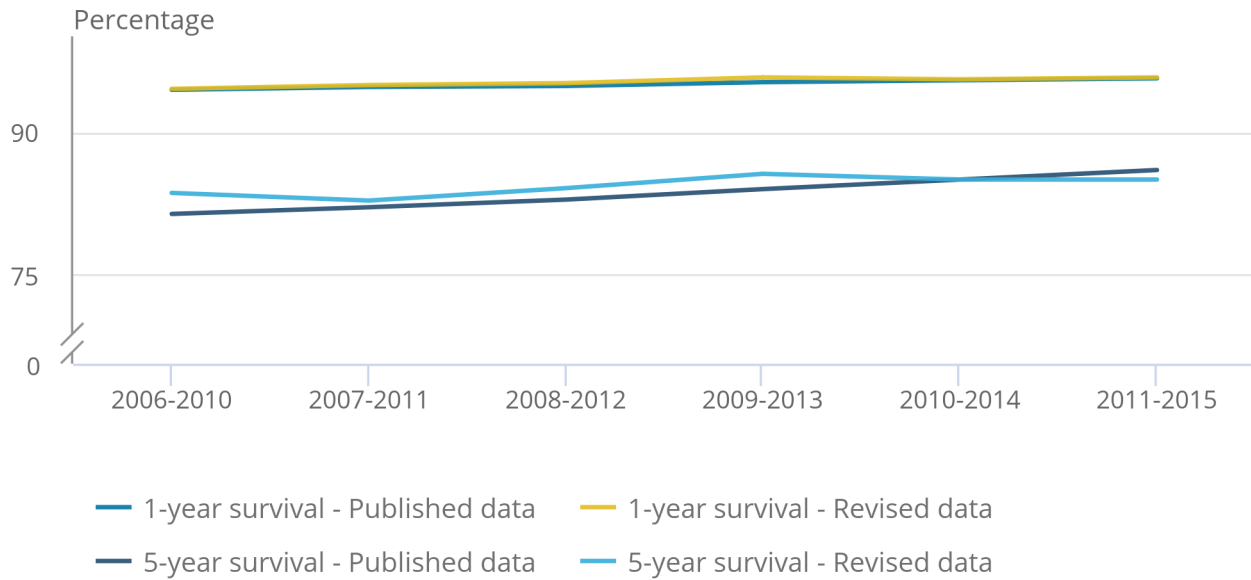
Compared with the previously published estimates, the adjustment to coding resulted in no impact (an average 0.0 percentage point change) across all cohorts. The cohort selection change resulted in a 0.1 percentage point increase on average. Changing the life table led to a 0.1 percentage point average increase in the estimates.

For 5-year survival, changes were generally larger than for 1-year survival. Survival estimates increased for each cohort up to 2009 to 2013, there was no change for 2010 to 2014 and a decrease for 2011 to 2015. The average change in 5-year survival was 0.8 percentage points, ranging from a 2.2 percentage point increase for 2006 to 2010 to a 1.0 percentage point decrease for 2011 to 2015. The changes were significant in all years apart from 2007 to 2011 and 2010 to 2014.

Compared with the previously published estimates, the adjustments to coding resulted in an average 0.6 percentage point decrease across all cohorts. The cohort selection change resulted in a 0.6 percentage point increase on average. Changing the life table led to a 0.9 percentage point average increase in the estimates.

Figure 1: Age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with breast cancer, England

Figure 1: Age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with breast cancer, England

Source: National Cancer Registration and Analysis Service within Public Health England; Office for National Statistics

As can be seen in Figure 1, the trend between 1-year published and revised data is very similar. Looking at 5-year survival, the revised data is more volatile cohort to cohort compared with published data. This is due to the fact there is less follow-up in the revised estimates compared with the published estimates. In the revised estimates, survival is slightly declining in the two most recent cohorts compared with increasing in the published data.

Impact of changes on prostate cancer

Table 2: Impact of changes on age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with prostate cancer, England

	1-year survival				5-year survival			
Cohort	Published data (%)	Revised data (%)	Difference	Significant change?	Published data (%)	Revised data (%)	Difference	Significant change?
2006-2010	94.8	94.6	-0.2	No	83.5	84.9	1.4	Yes
2007-2011	95.1	95.1	0.0	No	84.3	84.6	0.3	No
2008-2012	95.5	95.5	0.0	No	85.4	86.6	1.2	Yes
2009-2013	95.8	96.0	0.2	Yes	86.2	87.3	1.1	Yes
2010-2014	96.1	96.1	0.0	No	87.2	86.8	-0.4	No
2011-2015	96.3	96.3	0.0	No	88.3	86.6	-1.7	Yes

Source: Office for National Statistics and Public Health England

For 1-year survival, the majority of cohorts saw no change (0.0 percentage point differences) in survival estimates. In 2006 to 2010 the changes caused survival to decrease by 0.2 percentage points. The changes caused a 0.2 percentage point increase in 2009 to 2013, which was a significant increase.
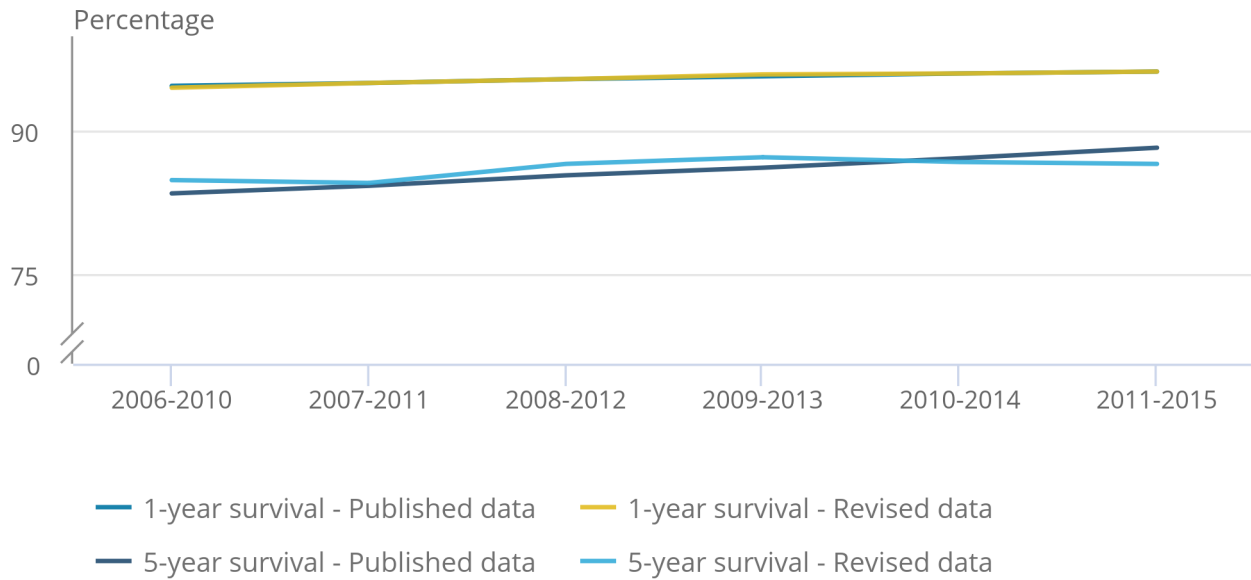
Compared with the previously published estimates, adjustments to coding had no impact (an average 0.0 percentage point change) across all cohorts. The cohort selection change resulted in a 0.1 percentage point decrease on average. Changing the life table led to a 0.0 percentage point average change in the estimates.

Similar to breast cancer, differences in 5-year survival were greater than that of 1-year survival, with an average difference of 0.3 percentage points. Survival increased between published and revised figures for the first four cohorts, all these increases were significant apart from 2007 to 2011. The survival estimate decreased between published and revised figures for 2010 to 2014 and 2011 to 2015, with 2011 to 2015 showing a significant decrease. Differences ranged from an increase of 1.4 percentage points in 2006 to 2010 to a decrease of 1.7 percentage points in 2011 to 2015.

Compared with the previously published estimates, the adjustments to coding resulted in an average 0.6 percentage point decrease across all cohorts. The cohort selection change resulted in a 0.6 percentage point increase on average. Changing the life table led to a 0.3 percentage point average increase in the estimates.

Figure 2: Age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with prostate cancer, England

Figure 2: Age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with prostate cancer, England

Source: National Cancer Registration and Analysis Service within Public Health England; Office for National Statistics

Looking at the trends over time, as with breast cancer, trends in 1-year published and revised estimates are similar. There is more volatility cohort to cohort in the revised 5-year survival estimates. Similar to breast cancer, prostate cancer shows a decline in the latest two cohorts in the revised estimates, whereas the published estimates show a slight increase.

Impact of changes on lung cancer

Table 3: Impact of changes on age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with lung cancer, England

Cohort	1-year survival				5-year survival			
	Published data (%)	Revised data (%)	Difference	Significant change?	Published data (%)	Revised data (%)	Difference	Significant change?
2006-2010	29.7	30.2	0.5	No	8.5	9.4	0.9	Yes
2007-2011	30.7	31.5	0.8	Yes	9.2	9.9	0.7	Yes
2008-2012	31.9	32.4	0.5	No	10.0	11.0	1.0	Yes
2009-2013	33.0	33.3	0.3	No	10.9	11.2	0.3	No
2010-2014	34.1	34.1	0.0	No	12.0	11.5	-0.5	No
2011-2015	35.4	35.3	-0.1	No	13.1	12.2	-0.9	Yes

Source: Office for National Statistics and Public Health England

For lung cancer in males, the 1-year survival estimate increased between the published and revised estimates for the first four cohorts, with 2007 to 2011 being a significant difference. 2010 to 2014 showed no difference and 2011 to 2015 showed a slight decrease. On average, survival estimates increased by 0.3 percentage points. This ranged from a 0.8 percentage point increase in 2007 to 2011 to a 0.1 percentage point decrease in 2011 to 2015.
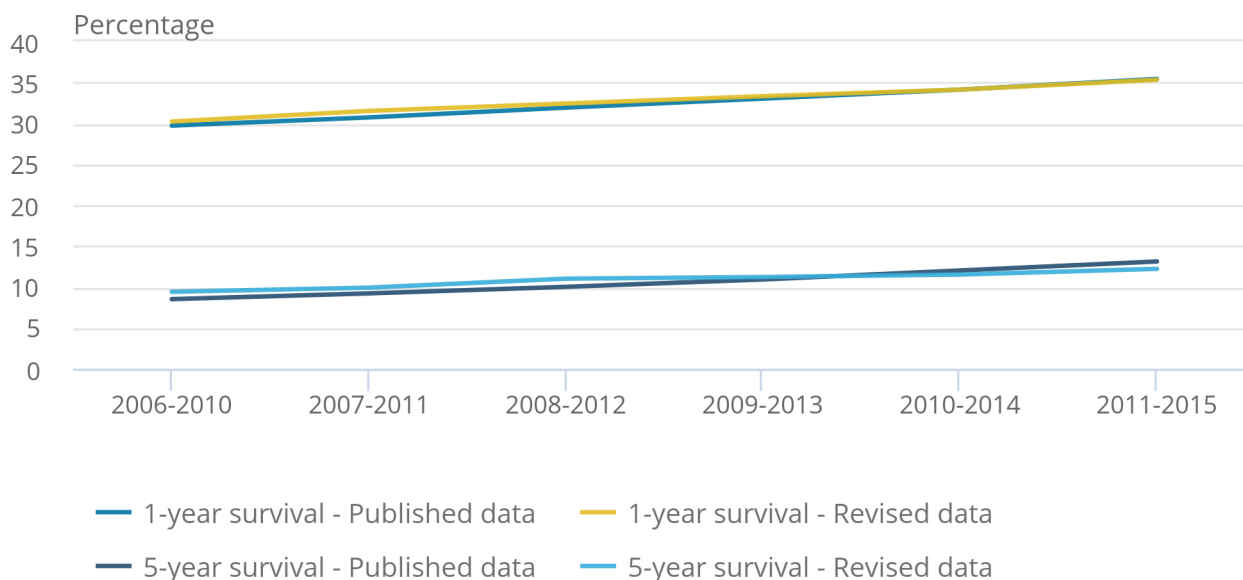
Compared with the previously published estimates, the adjustments to coding resulted in an average 0.0 percentage point change across all cohorts. The cohort selection change resulted in a 0.3 percentage point increase on average. Changing the life table led to a 0.0 percentage point average change in the estimates.

For 5-year survival, the revised estimates caused survival to increase between 2006 to 2010 and 2009 to 2013 with the differences being significant for the first three cohorts. The revised estimates caused survival estimates to decrease for 2010 to 2014 and 2011 to 2015, with the differences being significant in the last cohort. On average, survival increased by 0.2 percentage points. This ranged from a 1.0 percentage point increase in 2008 to 2012 to a 0.9 percentage point decrease in 2011 to 2015.

Compared with the previously published estimates, the adjustments to coding resulted in an average 0.3 percentage point decrease across all cohorts. The cohort selection change resulted in a 0.5 percentage point increase on average. Changing the life table led to a 0.0 percentage point average change in the estimates.

Figure 3: Age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with lung cancer, England

Figure 3: Age-standardised 1-year and 5-year net survival (%) for men (aged 15 to 99 years) diagnosed with lung cancer, England

Source: National Cancer Registration and Analysis Service within Public Health England; Office for National Statistics

The trends for both 1-year and 5-year survival estimates show a similar trend when looking at the published and revised estimates. In both the revised and published estimates, the 1-year and 5-year survival estimates increase cohort to cohort.

Table 4: Impact of changes on age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with lung cancer, England

| Cohort | 1-year survival | | | | 5-year survival | | | |
	Published data (%)	Revised data (%)	Difference	Significant change?	Published data (%)	Revised data (%)	Difference	Significant change?
2006-2010	33.7	33.9	0.2	No	11.0	11.6	0.6	No
2007-2011	35.2	35.7	0.5	No	12.0	12.4	0.4	No
2008-2012	37.1	37.4	0.3	No	13.3	14.2	0.9	Yes
2009-2013	38.9	39.0	0.1	No	14.7	14.9	0.2	No
2010-2014	40.5	40.4	-0.1	No	16.0	15.7	-0.3	No
2011-2015	42.0	41.9	-0.1	No	17.5	16.5	-1.0	Yes

Source: Office for National Statistics and Public Health England

For 1-year survival, the revised estimates showed an increase in survival between 2006 to 2010 and 2009 to 2013 and a decrease in survival for 2010 to 2014 and 2011 to 2015. None of these changes were significant. On average, the survival estimates increased by 0.2 percentage points. This ranged from an increase of 0.5 percentage points in 2007 to 2011 to a 1.0 percentage point decrease in 2011 to 2015.
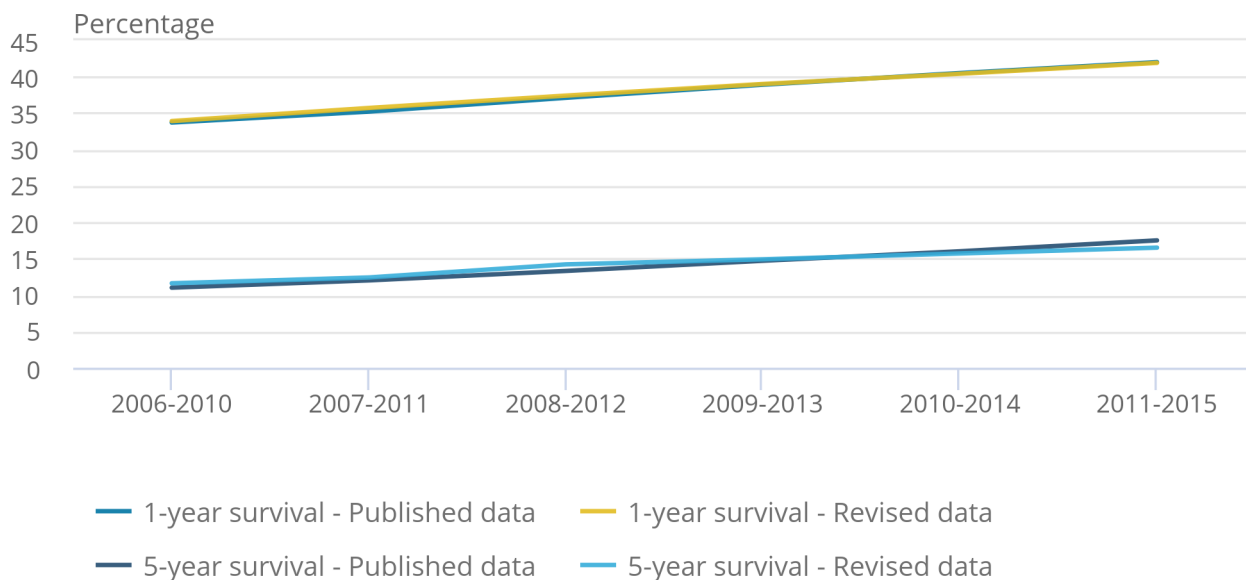
Compared with the previously published estimates, the adjustments to coding resulted in no change on average across all cohorts. The cohort selection change resulted in a 0.2 percentage point increase on average. Changing the life table led to no change on average in the estimates.

Comparing the published data with the revised data for 5-year survival shows a similar pattern to 1-year, with the revised estimates being higher than the published estimates for the first four cohorts and lower for the last two. On average, survival estimates increased by 0.1 percentage points. This ranged from an increase of 0.9 percentage points in 2008 to 2012 to a decrease of 1.0 percentage points in 2011 to 2015. The cohorts with significant differences were 2008 to 2012 and 2011 to 2015.

Compared with the previously published estimates, the adjustments to coding resulted in an average 0.3 percentage point decrease across all cohorts. The cohort selection change resulted in a 0.4 percentage point increase on average. Changing the life table led to a 0.1 percentage point average increase in the estimates.

Figure 4: Age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with lung cancer, England

Figure 4: Age-standardised 1-year and 5-year net survival (%) for women (aged 15 to 99 years) diagnosed with lung cancer, England

Source: National Cancer Registration and Analysis Service within Public Health England; Office for National Statistics

Similarly to lung cancer in males, the estimates for 1-year and 5-year survival show an increasing trend cohort-on-cohort for both published and revised estimates.

Summary of differences

Overall, different sites, age groups and genders will have been affected differently. The adjustment to coding has made small differences across all sites. The change to the cohort approach has affected different sites, age groups and genders randomly. The life tables have had more of an effect on sites with higher survival estimates and less of an effect on those with a low survival estimate, which is what we would expect as changes in mortality would have less of an effect on causes with high mortality.

The trend across cohorts has remained the same between published and revised data for lung cancer. In the revised estimates, the trend in survival for breast and prostate cancer now shows a decline for the most recent cohorts, compared with increasing in the published estimates. The same pattern can be seen across all cancer sites, with those sites with low survival estimates showing less of a change and sites with higher survival estimates showing more of a change. The reason for this is that there is a smaller proportion of patients alive in sites with low survival than in sites with high survival. Therefore the difference in the life tables can act on a smaller group of (still alive) patients in sites with low survival, which is why the poorer prognostic sites show less of an impact.

The comparisons show that there has been more of an effect on 5-year survival compared with 1-year survival. This is something we would expect as 5-year survival tends to be more volatile than 1-year survival, as the follow-up for 5-year survival is for a smaller proportion of the initial cohort (the first year of follow-up being complete for all diagnosis years but for fewer diagnosis years as the follow-up time increases). In aggregate, the changes reduce reported estimates of cancer survival for the two most recent periods of reporting (2010 to 2014 and 2011 to 2015). The reductions are larger for 5-year survival than 1-year survival and also for cancers with better prognosis than those with poor prognosis.

4 . Annex 1: Method for creating smoothed sub-national life tables

Rationale

The main purpose of producing sub-national life tables is to ensure the latest mortality and population trends are being applied in calculations of cancer survival analyses which consider background mortality. Background mortality is derived from population life tables.

To create cancer net survival, the mortality of cancer patients is compared with that of individuals in the general population who belong to the same single year of age (15 to 99 years), sex, quintile of the Index of Multiple Deprivation (IMD) and region. Patients are included if they were diagnosed with their tumour when they were between the ages of 15 and 99 years. Previously, the life tables used were supplied by the London School of Hygiene and Tropical Medicine's Cancer Survival Group.

Diagnoses in those aged 0 to 14 years are considered in the childhood cancer survival estimates. Generally, the consensus is that children with cancer will die from their cancer or complications of their cancer. Therefore, life tables are not needed for childhood cancer and there have been no revisions to the childhood cancer series.

Historically, the life tables calculated by the London School of Hygiene and Tropical Medicine (LSHTM) were created using census data. Due to recent changing mortality trends it became important to create life tables which use the most up to date mortality data available, to take into account the slowdown in the long-term improvements to mortality rates.

Office for National Statistics (ONS) currently produces life tables by single year of age at a national level and life tables by age group at a sub-national level. The life tables created for cancer survival will look at single year of age at a sub-national level. Data has been smoothed to overcome issues around high volatility and risk of disclosure.

These life tables have been developed with advice from University of Leicester. To quality assure the data, LSHTM created life tables with their methodology using the same data as ONS. Comparisons were then made between the two estimates and between survival estimates using both life tables.

Overview of life tables

There are different types of life tables which ONS currently publish. More information can be found in this explainer article.

Smoothed life tables

This paper includes an explanation on the method used to create a different type of lifetable – smoothed life tables. These smoothed life tables provide estimated (modelled) life expectancy by single year of age and sex over a 3-year period for each IMD quintile within each region of England – a total of 45 area estimates per year. Analysis is created over a 3-year period to smooth fluctuations due to exceptional events, for example, a "flu" epidemic. Smoothed life tables contain estimated mortality rates based on a statistical model, rather than rates calculated directly from the deaths registered and population estimates. The benefit of the smoothed life table approach is that it allows estimated complete life tables to be produced for areas whose population is too small for conventional complete life tables, for applications where abridged life tables are not appropriate.

The smoothed life tables are currently designated as Experimental Statistics. Experimental Statistics are published to involve customers and stakeholders in their development and as a means of building in quality at an early stage. Therefore, if you would like to provide feedback on these new life tables, please contact cancer.newport@ons.gov.uk.

Method for creating smoothed life tables

Data specification

Table 5 shows a breakdown of all the data included in the creation of smooth life tables. Populations and deaths data were aggregated into Index of Multiple Deprivation (IMD) quintiles. The Lower Super Output Areas (LSOAs) are weighted allocations, so there may not be an even number of LSOAs per quintile.

The life tables have been created for England only as the analysis is designed for the cancer survival estimates, which are only produced for England by ONS and Public Health England (PHE). Results would not be comparable across other UK countries since each country has its own IMD.

Table 5: Input data specification

Data	Notes
Populations and number of deaths by:	Mid-year population estimates and number of deaths registered
Sex	Males, Females
Age	Single year of age 0, 1, … 89, 90+
Rolling 3-year period	From "2001 to 2003" to "2014 to 2016" (as of October 2018)
Country	England
Regions	Regions of England (formally known as Government Office Regions (GOR)): North East, North West, Yorkshire and The Humber, South West, South East, East, East Midlands, West Midlands, London
Deprivation quintile	Index of Multiple Deprivation: 1 (most deprived), 2, 3, 4, 5 (least deprived) IMD over time: 2001 to 2005 used IMD 2004 2006 to 2008 used IMD 2007 2009 to 2012 used IMD 2010 2013 onwards used IMD 2015

Source: Office for National Statistics and Public Health England
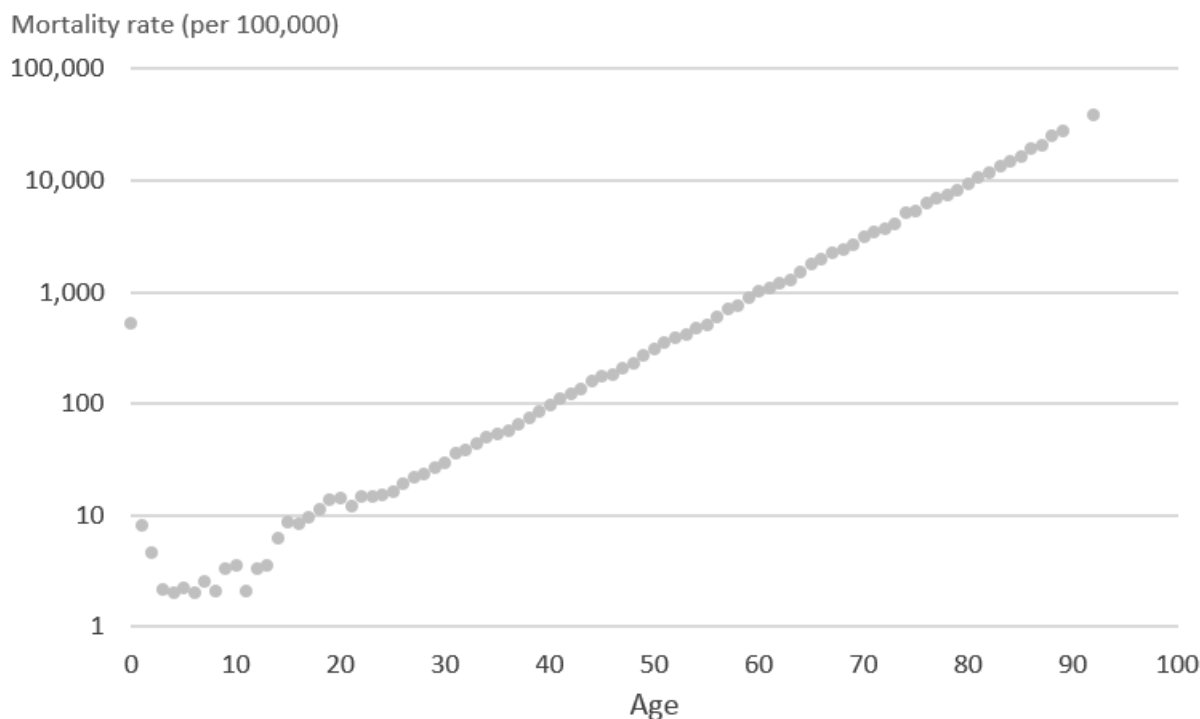
How we created smoothed life tables

The mortality data by age was smoothed to overcome issues around high volatility and high risk of disclosure for mortality at the youngest ages, and the need to extrapolate mortality rates for elderly. By smoothing mortality rates, we have been able to present smoothed patterns across ages at more granular level than was previously possible. This means that data can be produced by a region and IMD quintile level, which is needed for cancer survival estimates.

The method for creating smoothed life tables uses restricted cubic splines in an exponential Poisson regression. Restricted cubic splines are used when data has a non-linear trend that does not fit standard distributions. For example, standard distributions could include quadratic, cubic, exponential, logarithm, wave or circular distributions. Instead of trying to fit a line to all the data you could first break up the data into sections that follow smaller trends within the series and fit a line to each of these. Restricted cubic splines are a form of spline modelling, where spline functions are fitted between break points, also known as knots. This is demonstrated in Diagrams 1 to 3.

Cubic splines are piecewise functions, which are fitted between knots. Cubic splines use cubic polynomial functions that are forced to be continuous at the first and second derivative to smooth the joins at the knot placements. Restricted cubic splines means the cubic splines are forced to be linear beyond the first and last knots, known as boundary knots. We used fixed knots for age and the interaction between age and deprivation (that were selected a priori) for each regression. Normally we select knots at positions where we know that there is going to be a "change" in the direction of flow (that is, if the curve is known to continually increase but suddenly drops at around age 50 years, then we would select age 50 years as a position for a knot). Therefore, we can use the position of the knots to partition the mortality into different segments so that we are allowing flexibility for the mortality to fluctuate between age groups, without too many restrictions. Knots do not need to be equally spaced; for example, if we think that the variation in mortality is less volatile for age 15 to 45 years, we need not put a knot between these two ages just for the sake of it. Our selection of knots a priori is based on previous experience and knowledge gained from researchers in the past about the best positioning of knot locations. Different fixed knots were used for each sex due to the difference in mortality rates by age. The following diagrams show synthetic mortality rates on a logarithmic scale. The actual mortality rates that were smoothed have a high potential risk of disclosure so haven't been presented in this publication. However, to understand the smoothing technique, we created synthetic mortality rates that have no risk of disclosure but could be used to demonstrate how restricted cubic splines would smooth mortality data. These logarithmic scales start at 1 per 100,000 and increase 10-fold for each interval on the y-axis.

Diagram 1 shows synthetic mortality rates that have been generated to demonstrate how a restricted cubic spline can be fitted to mortality data. These age specific mortality rates have been plotted on a logarithmic scale to show the change in mortality rates seen by the young (aged under 25 years). By showing the graph on the logarithmic scale it is easier to interpret the effect of using restricted cubic splines.

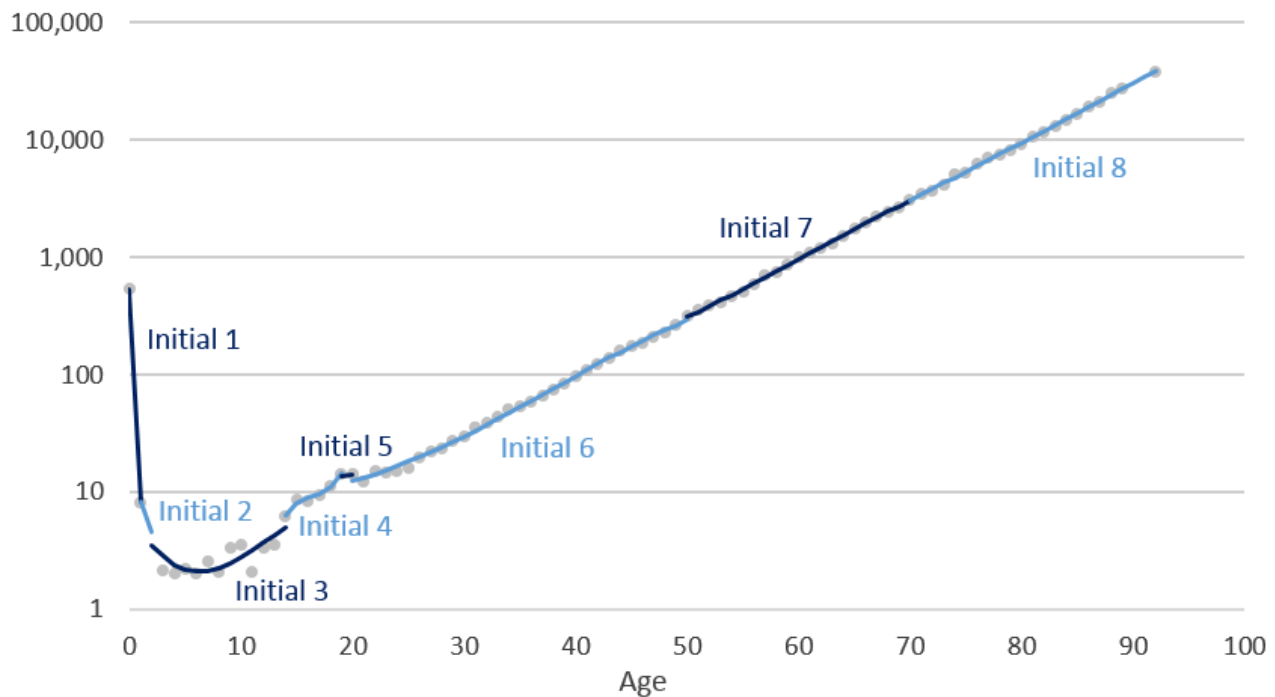**Diagram 1: Synthetic mortality rates (per 100,000 people) on the logarithmic scale**

Using the mortality rates in Diagram 1 we then applied the same knots at ages 0, 1, 2, 14, 19, 20, 50, 70 and 92 years, which are the knots used for male mortality rates, such that 0 and 92 form the boundary knots after which the model assumes a linear trajectory. Restricted cubic splines initially fit individual cubic polynomials between each pair of knots. Diagram 2 demonstrates what the initial cubic polynomials would look like for these mortality rates. Between each pair of knots (ages; 0 to 1, 1 to 2, 2 to 14, 14 to 19, 19 to 20, 20 to 50, 50 to 70, 70 to 92 years) initial cubic polynomials 1 to 8 have been fitted to the data.

Diagram 2: Synthetic mortality rates (per 100,000 people) and initial cubic polynomials on the logarithmic scale
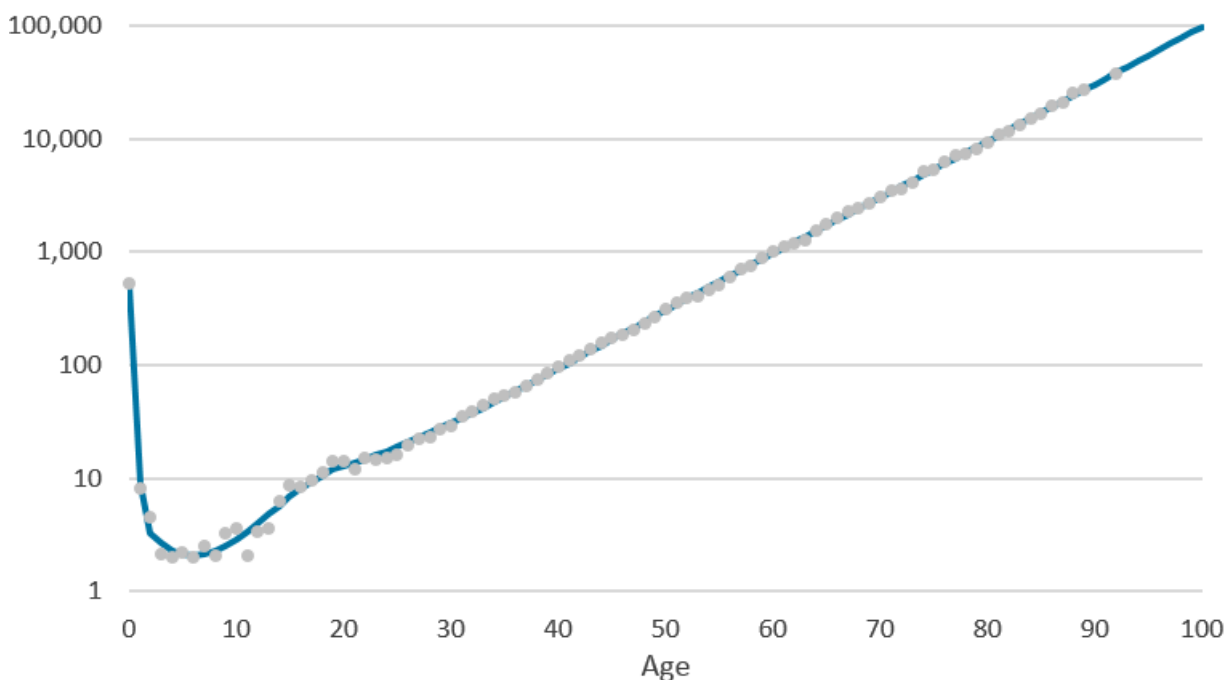

Source: Office for National Statistics

The piecewise polynomials are restricted so that two adjoining cubic splines are forced to meet at each knot. The final restricted cubic spline model is in Diagram 3 where all 8 cubic polynomials have been joined to form one smooth function that is linear beyond age 92 years, the boundary knot.

Diagram 3: Synthetic mortality rates (per 100,000 people) and final cubic spline on the logarithmic scale

Source: Office for National Statistics

In addition to using restricted cubic splines for the smooth life tables we applied a Poisson regression with a log link function and deprivation as a dependant variable, as this is the recommended methodology used by London School of Hygiene and Tropical Medicine (LSHTM). This approach was influenced by [Multivariable flexible modelling for estimating complete, smoothed life tables for sub-national populations](), which discusses four potential methods for calculating life tables and concludes that "The flexible Poisson model is recommended because it derives robust and unbiased estimates for sub-populations without making strong assumptions about age-specific mortality profiles. Fixing knots a priori in the final model greatly improves fit at the young ages". The flexible Poisson model uses mortality and populations data by single year of age, deprivation and the interaction between age and derivation. This model was then stratified by each calendar period, sex and region of England to give a total of 1260 life tables covering England between 2001 to 2003 and 2014 to 2016.

STATA 14 was used to estimate the smoothed mortality rates by applying the user written command rcsgen. The following fixed knots were used for all of the smoothed life tables. These were chosen because over several iterations these fixed knots had the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC). Age 90 years and over mortality rates were used to represent the mortality rate at the most common median age of death for males and females aged 90 years and over. The most common median age of death for males was 92 years and the most common for females was 93 years.

Table 6: Knots used for analysis

Sex	Knots for age	Knots for the interaction terms for age and deprivation
Males	0, 1, 2, 14, 19, 20, 50, 70, 92	0, 14, 32, 80, 92
Females	0, 1, 2, 14, 19, 50, 70, 93	0, 14, 32, 93

Source: Office for National Statistics and Public Health England

How we calculate life expectancy

Life expectancy calculations are the same as the [National life tables](), with the following exception. The mortality rate is based solely on the modelled mortality rates described above. Therefore, it makes no adjustment for infant mortality or use of populations estimates for the elderly. This differs from the national life tables which use population estimates calculated with the [Kannisto-Thatcher method](). Additionally, we only use period life expectancy for the smoothed life tables.

Strengths and weaknesses of smoothed life tables

Strengths:

- complete life tables can be published at a sub-national level without encountering errors in small numbers

- by using a robust smoothing technique, reliable estimated mortality rates for the young and elderly have been produced at a sub-national level.

- smooth life tables are based on data for a period of three consecutive years; the three-year rolling averages are used to smooth fluctuations due to exceptional events, for example, a "flu" epidemic

- timely life tables designed specifically for use in estimating cancer survival

Weaknesses:

- modelling the mortality rates reduces the amount that the rate represents the actual changes and variation in mortality experienced by the population, therefore, these smoothed rates are more useful for input into specific calculations such as cancer survival than for description of the individual geographical areas covered

- Different methods are used to produce the various life table products by ONS, therefore life expectancy figures are not comparable between different outputs; the national life tables should be used at national level, and the sub-national health expectancies figures based on abridged life tables are preferred for description of sub-national areas

- when modelling the mortality rates, we made assumptions about their distribution, these have been described more in the methods section

5 . Annex 2: Enhancement to the inclusion criteria – details of the changes

The coding system for the inclusion criteria of European Network of Cancer Registries' A proposal on cancer data quality checks: one common procedure for European cancer registries is specified in the International Classification of Diseases for Oncology, third edition (ICD-O3). This was in widespread use in England for diagnoses made in 2012 onwards.

Before then, England operated as a system of regional cancer registries and they adopted different coding systems which need to be converted into a ICD-O3 to enable the checks to be applied. The conversions have been made stricter by removing the possibility of implicit conversions from taking place.

These changes result in a small number of conversions being corrected (mainly applying to gynaecological and haematological cancers), increasing the number of tumours and patients that could be considered for affected rules.

There are a smaller number of conversions that are no longer covered in the stricter conversion sets that are no longer considered for survival analysis. The distribution of these tumours affected is spread evenly throughout the body and is not concentrated in any tumour site groupings.

6 . Acknowledgements

We would like to acknowledge the expertise and advice given to us by:

Paul Lambert and Mark Rutherford, University of Leicester, for assisting in the publication by reviewing the smoothing methodology.

Michel Coleman, Camille Maringe and Veronica Di Carlo, Cancer Survival Group, London School of Hygiene and Tropical Medicine, for quality assuring the proposed methodology for smoothed life tables.

John Broggio, Carolynn Gildea, Kwok Wong and Marta Emmett, Public Health England, for their work on the impact for the methodology changes on calculating net survival.